Linear Algebra and its Applications 436 (2012) 1960–1968

Contents lists available at SciVerse ScienceDirect

Linear Algebra and its Applications

journal homepage: www.elsevier.com/locate/laa

The sum of orthogonal matrices

Dennis I. Merino

Department of Mathematics, Southeastern Louisiana University, Hammond, LA 70402-0687, USA

ARTICLE INFO

Article history:
Received 14 October 2010
Accepted 3 October 2011
Available online 3 November 2011

Submitted by R.A. Brualdi

Keywords:
Sum of unitary matrices
Sum of orthogonal matrices

ABSTRACT

Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}, \mathbb{H}\}$. Let $\mathcal{U}_n(\mathbb{F})$ be the set of unitary matrices in $M_n(\mathbb{F})$, and let $\mathcal{O}_n(\mathbb{F})$ be the set of orthogonal matrices in $M_n(\mathbb{F})$. Suppose $n \geqslant 2$. We show that every $A \in M_n(\mathbb{F})$ can be written as a sum of matrices in $\mathcal{U}_n(\mathbb{F})$ and of matrices in $\mathcal{O}_n(\mathbb{F})$. Let $A \in M_n(\mathbb{F})$ be given and let $k \geqslant 2$ be the least integer that is a least upper bound of the singular values of A. When $\mathbb{F} = \mathbb{C}$, we show that A can be written as a sum of k matrices from $\mathcal{U}_n(\mathbb{F})$. When $\mathbb{F} = \mathbb{R}$, we show that if $k \leqslant 3$, then A can be written as a sum of 6 orthogonal matrices; if $k \geqslant 4$, we show that A can be written as a sum of $k+2$ orthogonal matrices.

1. Introduction

Let $\mathbb{F} = \mathbb{C}$ (the set of complex numbers) or $\mathbb{F} = \mathbb{R}$ (the set of real numbers). Let n be a given positive integer. We let $M_n(\mathbb{F})$ be the set of all n-by-n matrices with entries in $\mathbb{F}$. We also let $E_{ij} \in M_n(\mathbb{F})$ be the matrix whose (i, j) entry is 1 and all other entries are 0.

We study the sums of unitary matrices and we also study the sums of orthogonal matrices. We determine which matrices (if any) in $M_n(\mathbb{F})$ can be written as a sum of unitary or orthogonal matrices. We note that the sum of unitary matrices in $M_n(\mathbb{C})$ has been previously studied (see [5] and the references therein). Moreover, for $A, B \in M_n(\mathbb{C})$, sums of the form $UAU^* + VBV^*$, where $U, V \in M_n(\mathbb{C})$ are unitary, have also been studied [4].

We let $\mathcal{U}_n(\mathbb{F})$ be the set of unitary matrices in $M_n(\mathbb{F})$ and we let $\mathcal{O}_n(\mathbb{F})$ be the set of orthogonal matrices in $M_n(\mathbb{F})$.

We begin with the following observation.

Lemma 1. *Let n be a given positive integer. Let $\mathcal{G} \subset M_n(\mathbb{F})$ be a group under multiplication. Then $A \in M_n(\mathbb{F})$ can be written as a sum of matrices in $\mathcal{G}$ if and only if for every $Q, P \in \mathcal{G}$, the matrix QAP can be written as a sum of matrices in $\mathcal{G}$.*

E-mail address: dmerino@selu.edu

0024-3795/$ - see front matter
doi:10.1016/j.laa.2011.10.011

Notice that both $\mathcal{U}_n(\mathbb{F})$ and $\mathcal{O}_n(\mathbb{F})$ are groups under multiplication.

Let $\alpha \in \mathbb{F}$ be given. Then Lemma 1 guarantees that for each $Q \in \mathcal{G}$, we have that αQ can be written as a sum of matrices from $\mathcal{G}$ if and only if αI can be written as a sum of matrices from $\mathcal{G}$.

Lemma 2. *Let $n \geqslant 2$ be a given integer. Let $\mathcal{G} \subset M_n(\mathbb{F})$ be a group under multiplication. Suppose that $\mathcal{G}$ contains $K \equiv diag(1, -1, \ldots, -1)$ and the permutation matrices. Then every $A \in M_n(\mathbb{F})$ can be written as a sum of matrices in $\mathcal{G}$ if and only if for each $\alpha \in \mathbb{F}$, αI can be written as a sum of matrices in $\mathcal{G}$.*

Proof. The forward implication is trivial. For the other direction, suppose that for each $\alpha \in \mathbb{F}$, αI can be written as a sum of matrices in $\mathcal{G}$. Now, $K \in \mathcal{G}$ so that for each $\alpha \in \mathbb{F}$, Lemma 1 guarantees that αK can also be written as a sum of matrices in $\mathcal{G}$. It follows that $\alpha E_{11} = \frac{\alpha}{2} I + \frac{\alpha}{2} K$ can be written as a sum of matrices in $\mathcal{G}$. Now, for each $1 \leqslant i, j \leqslant n$, notice that $E_{ij} = PE_{11}Q$ for some permutation matrices P and Q, and that $P, Q \in \mathcal{G}$. Therefore, if $A \in M_n(\mathbb{F})$, then A can be written as a sum of matrices in $\mathcal{G}$, as $A = [a_{ij}] = \sum_{i,j} a_{ij} E_{ij}$. $\square$

2. Sum of orthogonal matrices

The only matrices in $\mathcal{O}_1(\mathbb{F})$ are ± 1. Hence, not every element of $M_1(\mathbb{F})$ can be written as a sum of elements in $\mathcal{O}_1(\mathbb{F})$. In fact, only the integers can be written as a sum of elements of $\mathcal{O}_1(\mathbb{F})$.

2.1. The case $\mathbb{F} = \mathbb{C}$

Notice that $\mathcal{U}_1(\mathbb{C}) = \left\{ e^{i\theta} : \theta \in \mathbb{R} \right\}$. Set $\mathcal{C}_2 \equiv \left\{ e^{i\theta} + e^{i\beta} : \theta, \beta \in \mathbb{R} \right\}$. If $\theta, \beta \in \mathbb{R}$ are given, then $\left| e^{i\theta} + e^{i\beta} \right| \leqslant 2$. Hence, $\mathcal{C}_2 \subset \mathcal{A}_2 \equiv \{z \in \mathbb{C} : |z| \leqslant 2\}$. We show that $\mathcal{A}_2 \subset \mathcal{C}_2$. Let $0 \leqslant r \leqslant 2$ be given. Set $\beta = -\theta$, and choose θ so that $2\cos\theta = r$. Then $e^{i\theta} + e^{-i\theta} = 2\cos\theta = r$. If $z = re^{i\delta}$, then choose $\beta = -\theta + 2\delta$, and choose θ so that $2\cos(\theta - \delta) = r$.

Let $k \geqslant 2$ be an integer. Set $\mathcal{C}_k \equiv \left\{ \sum_{j=1}^{k} e^{i\theta_j} : \theta_j \in \mathbb{R} \text{ for } j{=}1, \ldots, k \right\}$ and set $\mathcal{A}_k \equiv \{z \in \mathbb{C} : |z| \leqslant k\}$. We show that for each k, we have $\mathcal{A}_k = \mathcal{C}_k$.

First, notice that for each k, we have $\mathcal{C}_k \subset \mathcal{A}_k$. We now show that $\mathcal{A}_k \subset \mathcal{C}_k$. If $z = re^{i\beta}$, with $r, \beta \in \mathbb{R}$ and $r \geqslant 0$, then

$$e^{i\theta_1} + \cdots + e^{i\theta_k} = re^{i\beta} \text{ if and only if } e^{i(\theta_1 - \beta)} + \cdots + e^{i(\theta_k - \beta)} = r. \tag{1}$$

Hence, $z = re^{i\beta} \in \mathcal{C}_k$ if and only if $r \in \mathcal{C}_k$. For $\theta_1, \ldots, \theta_k \in \mathbb{R}$, set $f_k(\theta_1, \ldots, \theta_k) \equiv e^{i\theta_1} + \cdots + e^{i\theta_k}$.

The case $k = 2$ has already been shown. Let $k = 3$, and suppose $0 \leqslant r \leqslant 3$. Set $\theta_3 = 0$ and set $\theta_1 = \theta = -\theta_2$. Then, $f_3(\theta_1, \theta_2, \theta_3) = 1 + 2\cos\theta$, and θ may be chosen so that $0 \leqslant r \equiv 1 + 2\cos\theta \leqslant 3$.

We use mathematical induction to show the general case. The base cases $k = 2$ and $k = 3$ have already been shown. Assume that $k > 3$ and suppose that $\mathcal{C}_k = \mathcal{A}_k$.

Consider $f_{k+1}(\theta_1, \ldots, \theta_k, \theta_{k+1}) \equiv e^{i\theta_1} + \cdots + e^{i\theta_k} + e^{i\theta_{k+1}}$. Let $z = re^{i\beta}$ be given with $0 \leqslant r \leqslant k+1$. We show that $r \in \mathcal{C}_{k+1}$.

First, we show that $\mathcal{A}_2 \subseteq \mathcal{C}_{k+1}$. If k is even, choose $\theta_3 = \cdots = \theta_{k-1} = 0$ and $\theta_4 = \cdots = \theta_k = \pi$. Then $f_{k+1}(\theta_1, \ldots, \theta_k, \theta_{k+1}) = e^{i\theta_1} + e^{i\theta_2}$. If k is odd, choose $\theta_4 = \cdots = \theta_{k-1} = 0$ and $\theta_5 = \cdots = \theta_k = \pi$. Then $f_{k+1}(\theta_1, \ldots, \theta_k, \theta_{k+1}) \equiv e^{i\theta_1} + e^{i\theta_2} + e^{i\theta_3}$. In both cases, notice that $\mathcal{A}_2 \subseteq \mathcal{C}_{k+1}$. Hence, we may assume further that $r \geqslant 1$; that is, we need to show that $r \in \mathcal{C}_{k+1}$ for $1 \leqslant r \leqslant k+1$.

Choose $\theta_{k+1} = 0$, so that $f_{k+1}(\theta_1, \ldots, \theta_k, \theta_{k+1}) = f_k(\theta_1, \ldots, \theta_k) + 1$. Now, by our inductive hypothesis, the equation $f_k(\theta_1, \ldots, \theta_k) + 1 = r$ has a solution since $0 \leqslant r - 1 \leqslant k$.

Lemma 3. *Let $k \geqslant 2$ be a given integer. Let $\mathcal{A}_k \equiv \{z \in \mathbb{C} : |z| \leqslant k\}$ and let $\mathcal{C}_k \equiv \left\{ \sum_{j=1}^{k} e^{i\theta_j} : \theta_j \in \mathbb{R} \text{ for } j = 1, \ldots, k \right\}$. Then $\mathcal{A}_k = \mathcal{C}_k$.*

2.1.1. The case $\mathcal{U}_n(\mathbb{C})$

Let $\alpha \in \mathbb{C}$ be given. Then there exist an integer $k \geqslant 2$ and $\theta_1, \ldots, \theta_k \in \mathbb{R}$ such that $\alpha = f_k(\theta_1, \ldots, \theta_k)$. Now, notice that $\alpha I = f_k(\theta_1, \ldots, \theta_k) I = e^{i\theta_1} I + \cdots + e^{i\theta_k} I$ is a sum of matrices in $\mathcal{U}_n(\mathbb{C})$. When $n = 1$, every $\alpha \in \mathbb{C}$ can be written as a sum of elements of $\mathcal{U}_1(\mathbb{C})$. When $n \geqslant 2$, Lemma 2 guarantees that every $A \in M_n(\mathbb{C})$ can be written as a sum of matrices in $\mathcal{U}_n(\mathbb{C})$.

Lemma 4. *Let n be a given positive integer. Then every $A \in M_n(\mathbb{C})$ can be written as a sum of matrices in $\mathcal{U}_n(\mathbb{C})$.*

Let $A \in M_n(\mathbb{C})$ be given. We look at the number of matrices that make up the sum A.

Let $\alpha \in \mathbb{C}$ be given. If $|\alpha| \leqslant k$ for some positive integer k, then $\alpha \in \mathcal{A}_k$. Moreover, $\alpha \in \mathcal{A}_m$ for every integer $m \geqslant k$. For any such m, Lemma 3 guarantees that there exist $\theta_1, \ldots, \theta_m \in \mathbb{R}$ such that $\alpha = e^{i\theta_1} + \cdots + e^{i\theta_m}$. However, if $|\alpha| > k$, then $\alpha \notin \mathcal{A}_k$ and α cannot be written as a sum of k elements of $\mathcal{U}_1(\mathbb{C})$.

Write $A = U\Sigma V$ (the singular value decomposition of A, see [1, Theorem 7.3.5] or [2, Theorem 3.1.1]), where $U, V \in M_n(\mathbb{C})$ are unitary and $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_n)$ with $\sigma_1 \geqslant \cdots \geqslant \sigma_n \geqslant 0$. Let k be the least integer such that $\sigma_1 \leqslant k$. Suppose that $k \geqslant 2$. Then, for each l, we have $\sigma_l \in \mathcal{A}_k$. Moreover, $\sigma_1 \notin \mathcal{A}_{k-1}$. Hence, A cannot be written as a sum of $k - 1$ unitary matrices. However, for each l, we have $\sigma_l = e^{i\theta_{l1}} + \cdots + e^{i\theta_{lk}}$, where each $\theta_{l1}, \ldots, \theta_{lk} \in \mathbb{R}$. For each $t = 1, \ldots, k$, set $U_t = \mathrm{diag}\left(e^{i\theta_{1t}}, \ldots, e^{i\theta_{nt}}\right)$. Then $U_t \in M_n(\mathbb{C})$ is unitary and $\sum_{t=1}^{k} U_t = \Sigma$. Hence, A can be written as a sum of k unitary matrices.

Suppose that $k = 1$. If $\sigma_n = 1$, then $\Sigma = I$ and A is unitary. If $\sigma_n \neq 1$, then for each l, we have $\sigma_l \in \mathcal{A}_2$, and A can be written as a sum of two unitary matrices.

We summarize our results.

Theorem 5. *Let $A \in M_n(\mathbb{C})$ be given. Let k be the least (positive) integer so that there exist $U_1, \ldots, U_k \in \mathcal{U}_n(\mathbb{C})$ satisfying $U_1 + \cdots + U_k = A$.*

1. *If A is unitary, then $k = 1$.*
2. *If A is not unitary and $\sigma_1(A) \leqslant 2$, then $k = 2$.*
3. *If $m \geqslant 2$ is an integer such that $m < \sigma_1(A) \leqslant m + 1$, then $k = m + 1$.*

For positive integers $m \geqslant k$, we have $\mathcal{A}_k \subseteq \mathcal{A}_m$. Hence, every $U \in \mathcal{U}_n(\mathbb{C})$ can be written as a sum of two or more elements of $\mathcal{U}_n(\mathbb{C})$. It follows that every $A \in M_n(\mathbb{C})$ that can be written as a sum of k elements of $\mathcal{U}_n(\mathbb{C})$ can be written as a sum of m elements of $\mathcal{U}_n(\mathbb{C})$.

2.1.2. The case $\mathcal{O}_n(\mathbb{C})$

Let $n = 2$. Let $\alpha, \beta \in \mathbb{C}$ be given. Set

$$A(\alpha, \beta) \equiv \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}. \tag{2}$$

Choose β such that $\alpha^2 + \beta^2 = 1$ and notice that $A(\pm\alpha, \pm\beta) \in \mathcal{O}_2(\mathbb{C})$. Set $A_1 \equiv A(\alpha, \beta)$ and set $A_2 \equiv A(\alpha, -\beta)$. Then $A_1 + A_2 = 2\alpha I_2$. Lemma 2 guarantees that every $A \in M_2(\mathbb{C})$ can be written as a sum of matrices from $\mathcal{O}_2(\mathbb{C})$.

We look at the case when $n = 3$. Let $\alpha, \beta \in \mathbb{F}$ be given. Set

$$B(\alpha, \beta) \equiv \begin{bmatrix} 1 & 0 & 0 \\ 0 & \alpha & \beta \\ 0 & -\beta & \alpha \end{bmatrix}, \tag{3}$$

set

$$C(\alpha,\beta) \equiv \begin{bmatrix} \alpha & 0 & \beta \\ 0 & 1 & 0 \\ -\beta & 0 & \alpha \end{bmatrix}, \tag{4}$$

and set

$$D(\alpha,\beta) \equiv \begin{bmatrix} \alpha & \beta & 0 \\ -\beta & \alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{5}$$

Choose β so that $\alpha^2+\beta^2=1$. Then, $B(\pm\alpha,\pm\beta)$, $C(\pm\alpha,\pm\beta)$, and $D(\pm\alpha,\pm\beta)$ are all elements of $\mathcal{O}_3(\mathbb{C})$. Set $B_1 \equiv B(\alpha,\beta)$, set $B_2 \equiv B(-\alpha,\beta)$, set $C_1 \equiv C(\alpha,\beta)$, set $C_2 \equiv C(-\alpha,\beta)$, set $D_1 \equiv D(\alpha,\beta)$, and set $D_2 \equiv D(-\alpha,\beta)$. Then, $B_1-B_2+C_1-C_2+D_1-D_2=2\alpha I_3$. Lemma 2 now guarantees that every $A \in M_3(\mathbb{C})$ can be written as a sum of matrices in $\mathcal{O}_3(\mathbb{C})$.

Let $n=2m$ be a positive even integer, and let $\delta \in \mathbb{C}$ be given. Choose $A_1, A_2 \in \mathcal{O}_2(\mathbb{C})$ so that $A_1+A_2=\delta I_2$. Set $E_1 = A_1 \oplus \cdots \oplus A_1$ (m copies) and set $E_2 = A_2 \oplus \cdots \oplus A_2$ (m copies). Then $E_1, E_2 \in \mathcal{O}_{2m}(\mathbb{C})$, and $E_1+E_2=\delta I_{2m}$.

Let $n=2m+1 \geqslant 3$ be an odd integer, and let $\delta \in \mathbb{C}$ be given. Choose $A_1, A_2 \in \mathcal{O}_2(\mathbb{C})$ so that $A_1+A_2=\delta I_2$. Also, choose $B_1, B_2, C_1, C_2, D_1, D_2 \in \mathcal{O}_3(\mathbb{C})$ such that $B_1-B_2+C_1-C_2+D_1-D_2=\delta I_3$. Set $E_1 = A_1 \oplus \cdots \oplus A_1 \oplus B_1$ ($m-1$ copies of A_1), set $E_2 = A_2 \oplus \cdots \oplus A_2 \oplus -B_2$ ($m-1$ copies of A_2), set $E_3 = I_{2m-2} \oplus C_1$, set $E_4 = -I_{2m-2} \oplus -C_2$, set $E_5 = I_{2m-2} \oplus D_1$, and set $E_6 = -I_{2m-2} \oplus -D_2$. Then each $E_j \in \mathcal{O}_{2m+1}(\mathbb{C})$, and $E_1+\cdots+E_6=\delta I_{2m+1}$.

Hence, for every $\alpha \in \mathbb{C}$ and for every integer $n \geqslant 2$, αI can be written as a sum of matrices from $\mathcal{O}_n(\mathbb{C})$. Lemma 4 guarantees that every $A \in M_n(\mathbb{C})$ can be written as a sum of matrices from $\mathcal{O}_n(\mathbb{C})$.

Theorem 6. *Let $n \geqslant 2$ be a given integer. Then every $A \in M_n(\mathbb{C})$ can be written as a sum of matrices from $\mathcal{O}_n(\mathbb{C})$.*

Suppose that $A = Q_1 + Q_2$, where $Q_1, Q_2 \in \mathcal{O}_n(\mathbb{C})$. Then one checks that $AA^T = Q_1A^TAQ_1^T$, so that AA^T and A^TA are similar. Theorem 13 of [3] ensures that $A = QS$, where Q is orthogonal and S is symmetric (or that A has a QS decomposition). Suppose now that has a QS decomposition. Is it true that A can be written as a sum of two (complex) orthogonal matrices? Take the case $n=1$, and notice that every $A \in M_n(\mathbb{C})$ is a scalar and has a QS decomposition. However, only the integers can be written as a sum of orthogonal matrices in this case.

Lemma 7. *Let an integer $n \geqslant 2$ and $0 \neq \alpha \in \mathbb{C}$ be given. If $\alpha I = Q + V$ is a sum of two matrices from $\mathcal{O}_n(\mathbb{C})$, then there exists a skew-symmetric $D \in M_n(\mathbb{C})$ such that $Q = \frac{\alpha}{2}I + D$, $V = \frac{\alpha}{2}I - D$, and $DD^T = \left(1-\frac{\alpha^2}{4}\right)I$. Conversely, if there exists a skew-symmetric $D \in M_n(\mathbb{C})$ such that $DD^T = \left(1-\frac{\alpha^2}{4}\right)I$, then $Q \equiv \frac{\alpha}{2}I + D$ and $V \equiv \frac{\alpha}{2}I - D$ are in $\mathcal{O}_n(\mathbb{C})$ and $Q+V=\alpha I$.*

Proof. Let an integer $n \geqslant 2$ and $0 \neq \alpha \in \mathbb{C}$ be given. Suppose that $\alpha I \in M_n(\mathbb{C})$ can be written as a sum of two orthogonal matrices, say, $\alpha I = Q + V$. Write $Q = [a_{ij}] = [q_1 \dots q_n]$ and $V = [b_{ij}] = [v_1 \dots v_n]$. Then, $b_{ij} = -a_{ij}$ for $i \neq j$. Now, for each $i = 1, \dots, n$, we have $\sum_{j=1}^n a_{ij}^2 = q_i^Tq_i = 1 = v_i^Tv_i = \sum_{j=1}^n b_{ij}^2 = b_{ii}^2 + \sum_{j\neq i, j=1}^n a_{ij}^2$. Hence, $b_{ii} = \pm a_{ii}$. Because $Q+V=\alpha I$ and $\alpha \neq 0$, we have $b_{ii} = a_{ii} = \frac{\alpha}{2}$. Set $D = [d_{ij}]$, with $d_{ij} = a_{ij}$ if $i \neq j$, and $d_{ii} = 0$, so that $Q = \frac{\alpha}{2}I + D$ and $V = \frac{\alpha}{2}I - D$.

Now, since Q and V are orthogonal, we have

$$QQ^T = \frac{\alpha^2}{4}I + \frac{\alpha}{2}\left(D + D^T\right) + DD^T = I \tag{6}$$

and

$$VV^T = \frac{\alpha^2}{4}I - \frac{\alpha}{2}\left(D + D^T\right) + DD^T = I. \tag{7}$$

Subtracting Eq. (7) from Eq. (6), we get $D = -D^T$, so that D is skew-symmetric. Moreover, $DD^T = \left(1 - \frac{\alpha^2}{4}\right) I$.

For the converse, suppose that $D \in M_n(\mathbb{C})$ is skew-symmetric and satisfies $DD^T = \left(1 - \frac{\alpha^2}{4}\right) I$. Set $Q \equiv \frac{\alpha}{2}I + D$ and set $V \equiv \frac{\alpha}{2}I - D$. Then one checks that both Q and V are orthogonal matrices and $Q + V = \alpha I$. $\square$

If $\alpha = 0$, then for any orthogonal Q, notice that $\alpha I = Q + (-Q)$ is a sum of two orthogonal matrices. Let $n = 2$ and $\alpha \neq 0$. Set $\beta \equiv \sqrt{1 - \frac{\alpha^2}{4}}$ (either square root). Then $D \equiv \begin{bmatrix} 0 & \beta \\ -\beta & 0 \end{bmatrix}$ is skew-symmetric and satisfies $DD^T = \left(1 - \frac{\alpha^2}{4}\right) I$. Lemma 7 guarantees that αI can be written as a sum of two orthogonal matrices. If $n = 2k$ and $\alpha \neq 0$, set $E \equiv D \oplus \cdots \oplus D$ (k copies) and notice that E is skew-symmetric and satisfies $EE^T = \left(1 - \frac{\alpha^2}{4}\right) I$. Hence, if $n = 2k$ and if α is a scalar, then αI can be written as a sum of two orthogonal matrices.

Theorem 8. *Let n be a given positive integer. For each $\alpha \in \mathbb{C}$ and each orthogonal $Q \in M_{2n}(\mathbb{C})$, αQ can be written as a sum of two orthogonal matrices.*

Let an integer $n \geqslant 2$ be given. If $\alpha \in \{-2, 0, 2\}$, then one checks that $\alpha I \in M_n(\mathbb{C})$ can be written as a sum of two orthogonal matrices.

Theorem 9. *Let $\alpha \in \mathbb{C}$ and let a positive integer n be given. Then $\alpha I \in M_{2n+1}(\mathbb{C})$ can be written as a sum of two matrices from $\mathcal{O}_n(\mathbb{C})$ if and only if $\alpha \in \{-2, 0, 2\}$.*

Proof. For the forward implication, let $\alpha \in \mathbb{C}$ and let a positive integer n be given. Suppose that $\alpha I \in M_{2n+1}(\mathbb{C})$ can be written as a sum of two orthogonal matrices. Then $\alpha = 0$ or $\alpha \neq 0$. If $\alpha = 0$, then $\alpha \in \{-2, 0, 2\}$. If $\alpha \neq 0$, we show that $\alpha = \pm 2$. Lemma 7 guarantees that there exists a skew-symmetric $D \in M_n(\mathbb{C})$ satisfying $DD^T = \left(1 - \frac{\alpha^2}{4}\right) I$. Now, since n is odd, the skew-symmetric D is singular. Hence, DD^T is singular and $\alpha = \pm 2$.

The backward implication can be shown by direct computation. $\square$

2.2. The case $\mathbb{F} = \mathbb{R}$

Notice that $\mathcal{O}_n(\mathbb{R}) = \mathcal{U}_n(\mathbb{R})$. When $n = 1$, only the integers can be written as a sum of elements of $\mathcal{O}_1(\mathbb{R})$. Suppose that $n = 2$. We mimic the computations done in the case when $\mathbb{F} = \mathbb{C}$. Let $\theta \in \mathbb{R}$ be given, set $\alpha = \cos\theta$, and set $\beta = \sin\theta$. Then $A(\alpha, \beta)$ in Eq. (2) is an element of $\mathcal{O}_2(\mathbb{R})$. Moreover, $A_1 + A_2 = 2\cos\theta I_2$. Now, for every $\delta \in \mathbb{R}$, there exist a positive integer m and $\theta \in \mathbb{R}$ such that $2m\cos\theta = \delta$. We conclude that every $A \in M_2(\mathbb{R})$ can be written as a sum of an even number of matrices from $\mathcal{O}_2(\mathbb{R})$.

When $n = 3$, we again mimic the computations done in the case when $\mathbb{F} = \mathbb{C}$ using $\alpha = \cos\theta$ and $\beta = \sin\theta$ to show that for every $\delta \in \mathbb{R}$, the matrix δI_3 can be written as a sum of an even number of matrices from $\mathcal{O}_3(\mathbb{R})$.

Let $n \geqslant 4$ be a given integer. If $n = 2k$ is even, then write $\delta I_{2k} = \delta I_2 \oplus \cdots \oplus \delta I_2$ (k copies), and note that each δI_2 can be written as a sum of an even number of matrices from $\mathcal{O}_2(\mathbb{R})$. If $n = 2k + 1$ is odd, then write $\delta I_{2k+1} = \delta I_{2n-2} \oplus \delta I_3$. Now, δI_{2n-2} can be written as a sum of an even number of matrices from $\mathcal{O}_{2n-2}(\mathbb{R})$ and δI_3 can be written as a sum of an even number of matrices from $\mathcal{O}_3(\mathbb{R})$. We conclude that δI_{2k+1} can be written as a sum of an even number of matrices from $\mathcal{O}_{2k+1}(\mathbb{R})$.

Hence, Lemma 4 guarantees that for every integer $n \geqslant 2$, every $A \in M_n(\mathbb{R})$ can be written as a sum of matrices from $\mathcal{O}_n(\mathbb{R})$.

Theorem 10. *Let $n \geqslant 2$ be a given integer. Every $A \in M_n(\mathbb{R})$ can be written as a sum of matrices from $\mathcal{O}_n(\mathbb{R}) = \mathcal{U}_n(\mathbb{R})$.*

Let $n \geqslant 2$ be a given integer and let $U \in \mathcal{U}_n(\mathbb{R})$ be given. Then $U \in \mathcal{U}_n(\mathbb{C}) \cap \mathcal{O}_n(\mathbb{C})$, that is, a real orthogonal matrix is both complex unitary and complex orthogonal. Hence, an $A \in M_n(\mathbb{R})$, which is a sum of matrices in $\mathcal{U}_n(\mathbb{R})$ is both a sum of complex unitary matrices and a sum of complex orthogonal matrices. Thus, the restrictions on these cases apply. If k is a positive integer such that $\sigma_1(A) > k$, then A cannot be written as a sum of k real orthogonal matrices.

Let m be a positive integer. Theorem 9 guarantees that $I \in M_{2m+1}(\mathbb{C})$ cannot be written as a sum of two matrices in $\mathcal{O}_{2m+1}(\mathbb{C})$. Now, I cannot be written as a sum of two matrices from $\mathcal{O}_{2m+1}(\mathbb{R}) \subset \mathcal{O}_{2m+1}(\mathbb{C})$. In general, if $\alpha \notin \{-2, 0, 2\}$ and if $Q \in \mathcal{O}_{2m+1}(\mathbb{R})$, then αQ cannot be written as a sum of two matrices from $\mathcal{O}_{2m+1}(\mathbb{R})$.

Let $n \geqslant 2$ be a given integer, and let $A \in M_n(\mathbb{R})$ be given. We now look at the matrices in $\mathcal{O}_n(\mathbb{R})$ that make up the sum A.

Definition 11. Let $\theta \in \mathbb{R}$ be given. We define

$$A(\theta) \equiv \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \text{ and } B(\theta) \equiv \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}. \tag{8}$$

Set $K_2 \equiv B(0)$ and notice that $A(0) = I_2$. Let $0 \leqslant r, s \in \mathbb{R}$ be given, and let $k \geqslant 2$ be an integer. If $r, s \leqslant k$, then Lemma 3 and taking the real and imaginary parts of the equation

$$e^{i\theta_1} + \cdots + e^{i\theta_k} = \alpha \tag{9}$$

show that there exist $\theta_1, \ldots, \theta_k \in \mathbb{R}$ such that $A(\theta_1) + \cdots + A(\theta_k) = rI_2$. Moreover, there exist $\beta_1, \ldots, \beta_k \in \mathbb{R}$ such that $B(\beta_1) + \cdots + B(\beta_k) = sK_2$.

Theorem 12. *Let a positive integer n and let $A \in M_{2n}(\mathbb{R})$ be given. Suppose that $k \geqslant 2$ is an integer such that $\sigma_1(A) \leqslant k$. Then A can be written as a sum of $2k$ matrices in $\mathcal{O}_{2n}(\mathbb{R})$. Moreover, for every integer $m \geqslant 2k$, the matrix A can be written as a sum of m matrices in $\mathcal{O}_{2n}(\mathbb{R})$.*

Proof. Let $A = U\Sigma V$ be a singular value decomposition of A. Then Lemma 1 guarantees that we only need to concern ourselves with Σ.

For $n = 1$, notice that $\mathrm{diag}(\sigma_1, \sigma_2) = sI_2 + tK_2$, where $s = \frac{1}{2}(\sigma_1 + \sigma_2)$ and $t = \frac{1}{2}(\sigma_1 - \sigma_2)$. Now, $0 \leqslant t \leqslant s \leqslant k$. Notice that sI_2 and tK_2 can each be written as a sum of k orthogonal matrices. Moreover, for each integer $p \geqslant k$, notice that sI_2 can be written as a sum of p orthogonal matrices. Hence, $sI_2 + tK_2$ can be written as a sum of $p + k$ orthogonal matrices.

For $n > 1$, write $\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_{2n-1}, \sigma_{2n}) = \mathrm{diag}(\sigma_1, \sigma_2) \oplus \cdots \oplus \mathrm{diag}(\sigma_{2n-1}, \sigma_{2n})$. Notice now that for each $j = 1, \ldots, n$, $\mathrm{diag}(\sigma_{2j-1}, \sigma_{2j})$ can be written as a sum of $2k$ orthogonal matrices, say $P_{j1}, \ldots, P_{j(2k)}$. For each $l = 1, \ldots, 2k$, set $Q_l \equiv P_{1l} \oplus \cdots \oplus P_{nl}$, and notice that $\Sigma = Q_1 + \cdots + Q_{2k}$.

Finally, notice that for each integer $m \geqslant 2k$ and for each $j = 1, \ldots, n$, the matrix $\mathrm{diag}(\sigma_{2j-1}, \sigma_{2j})$ is also a sum of m orthogonal matrices. $\square$

Consider $C_0 \equiv \mathrm{diag}(b, a)$ with real numbers $b \geqslant a \geqslant 0$. If $b \leqslant 2$, then Theorem 12 ensures that C_0 can be written as a sum of 4 real orthogonal matrices. Moreover, for each integer $t \geqslant 4$, C_0 can be written as a sum of t real orthogonal matrices.

Suppose that $b \leqslant 3$. If $0 \leqslant b \leqslant 2$, then Theorem 12 guarantees that C_0 can be written as a sum of 4 real orthogonal matrices. Moreover, C_0 can also be written as a sum of 5 real orthogonal matrices. If $2 < b \leqslant 3$, then we look at two cases: (i) $0 \leqslant a \leqslant 1$ and (ii) $1 \leqslant a \leqslant 3$. In the first case, set $C_1 \equiv C_0 - K_2$. Then $0 \leqslant b - 1 \leqslant 2$ and $0 \leqslant a + 1 < 2$. Notice now that for each integer $t \geqslant 4$, C_1

can be written as a sum of t real orthogonal matrices. In the second case, set $C_1 \equiv C_0 - I_2$. Then we have $0 \leqslant a - 1 \leqslant b - 1 \leqslant 2$. Theorem 12 guarantees that for each integer $t \geqslant 4$, C_1 can be written as a sum of t real orthogonal matrices. Hence, for each integer $t \geqslant 5$, C_0 can be written as a sum of t real orthogonal matrices.

We now use induction to show that if $k \geqslant 2$ is an integer satisfying $b \leqslant k$, then for each integer $t \geqslant k + 2$, C_0 can be written as a sum of t real orthogonal matrices.

Suppose that the claim is true for some integer $k \geqslant 3$. We show that the claim is true when $0 \leqslant b \leqslant k + 1$. If $0 \leqslant b \leqslant k$, then our inductive hypothesis guarantees that for each integer $t \geqslant k + 2$, C_0 can be written as a sum of t (and hence, also of $t \geqslant k + 3$) real orthogonal matrices. If $k < b \leqslant k + 1$, we take a look at two cases: (i) $1 \leqslant a \leqslant k + 1$ and (ii) $0 \leqslant a < 1$. In case (i), set $C_1 \equiv C_0 - I_2$; and in case (ii), set $C_1 \equiv C_0 - K_2$.

We summarize our results.

Lemma 13. *Let $C \in M_2(\mathbb{R})$ be given. Suppose that $k \geqslant 2$ is an integer such that $\sigma_1(C) \leqslant k$. Then for each integer $t \geqslant k + 2$, C can be written as a sum of t matrices from $\mathcal{U}_2(\mathbb{R})$.*

Let $A \in M_{2n}(\mathbb{R})$ be given, and suppose that the singular values of A are $\sigma_1 \geqslant \cdots \geqslant \sigma_{2n} \geqslant 0$. Set $D \equiv \mathrm{diag}(\sigma_1, \ldots, \sigma_{2n})$. Write $D = \mathrm{diag}(\sigma_1, \sigma_2) \oplus \cdots \oplus \mathrm{diag}(\sigma_{2n-1}, \sigma_{2n})$. Let $k \geqslant 2$ be an integer such that $\sigma_1(A) \leqslant k$. Then Lemma 13 guarantees that for each integer $t \geqslant k+2$, and for each $j = 1, \ldots, n$, $\mathrm{diag}(\sigma_{2j-1}, \sigma_j)$ can be written as a sum of t real orthogonal matrices. We conclude that for each integer $t \geqslant k + 2$, A can be written as a sum of t real orthogonal matrices.

Theorem 14. *Let n be positive integer, and let $A \in M_{2n}(\mathbb{R})$ be given. Suppose that $k \geqslant 2$ is an integer such that $\sigma_1(A) \leqslant k$. Then for each integer $t \geqslant k + 2$, A can be written as a sum of t matrices in $\mathcal{U}_{2n}(\mathbb{R})$.*

Let $A \in M_3(\mathbb{R})$ be given. Suppose that $A = P\Sigma Q$, with $P, Q \in \mathcal{O}_3(\mathbb{R})$ and $\Sigma = \mathrm{diag}(a, b, c)$ with $0 \leqslant c \leqslant b \leqslant a \leqslant 2$. If $a = 2$, then notice that $\mathrm{diag}(b, c)$ can be written as a sum of four orthogonal matrices. One checks that Σ can be written as a sum of four real orthogonal matrices. Suppose $a < 2$. If $c = 0$, then Σ can be written as a sum of four orthogonal matrices. If $0 \neq c < 2$, then choose θ so that $2\cos\theta = c$. Notice that $A(\theta) + A(-\theta) = 2\cos\theta I_2$. Set $U_1 = [1] \oplus A(\theta)$ and set $U_2 = [-1] \oplus A(-\theta)$. Then $\Sigma - (U_1 + U_2) = \mathrm{diag}(a, b - c, 0)$, which can be written as a sum of four real orthogonal matrices. Hence, A can be written as a sum of six real orthogonal matrices.

Let n be a given positive integer and let $A \in M_{2n+1}(\mathbb{R})$. Let $A{=}P\Sigma Q$ be the singular value decomposition of A. Suppose that k is a positive integer such that $\sigma_1(A) \leqslant k$. If $k \leqslant 2$, then A can be written as a sum of at most six matrices in $\mathcal{O}_{2n+1}(\mathbb{R})$. If $k > 2$ and if $\sigma_1(A) = k$, we write $\Sigma = [k] \oplus \mathrm{diag}(\sigma_2, \sigma_3) \oplus \cdots \oplus \mathrm{diag}(\sigma_{2n}, \sigma_{2n+1})$. Notice that each of the 2-by-2 matrices can be written as a sum of $k{+}2$ (or more) orthogonal matrices. Hence, Σ can be written as a sum of $k{+}2$ matrices from $\mathcal{O}_{2n+1}(\mathbb{R})$.

Suppose that $\sigma_1 < k$. If $k = 3$, choose θ so that $\sigma_1 - 2\cos\theta = 2$. For $j = 3, \ldots, 2n+1$, set $d_j = 1$; if $\sigma_j \geqslant 2$, then set $e_j = 1$, and if $\sigma_j < 2$, then set $e_j = -1$. Set $D = \mathrm{diag}(d_3, \ldots, d_{2n+1})$ and set $E = \mathrm{diag}(e_3, \ldots, e_{2n+1})$. Set $U_1 = A(\theta) \oplus D$ and set $U_2 = A(-\theta) \oplus E$. Notice that $\Sigma - (U_1 + U_2)$ can be written as a sum of four orthogonal matrices, so that Σ is a sum of six orthogonal matrices.

If $k \geqslant 4$, choose θ so that $\sigma_1 - 2\cos\theta = k - 2$. Make the same choices for D and E as in the case $k = 3$, and also make the same choices for U_1 and U_2. Notice now that Σ can be written as a sum of $k + 2$ orthogonal matrices.

We summarize our results.

Theorem 15. *Let $A \in M_{2n+1}(\mathbb{R})$ be given. Suppose $k \geqslant 2$ is an integer such that $\sigma_1(A) \leqslant k$. If $k \leqslant 3$, then A can be written as a sum of at most six matrices in $\mathcal{O}_{2n+1}(\mathbb{R})$. If $k \geqslant 4$, then A can be written as a sum of $k + 2$ matrices in $\mathcal{O}_{2n+1}(\mathbb{R})$.*

Let $A \in M_{2n+1}(\mathbb{R})$ be given. Write $A = P\Sigma Q$, with the singular values arranged in decreasing order, that is, $\sigma_1 \geqslant \cdots \geqslant \sigma_{2n+1}$. Let $U = \mathrm{diag}(u_1, \ldots, u_{2n+1})$, where $u_i = 1$ if $\sigma_i \geqslant 1$ and $u_i = -1$ otherwise. Consider $\Sigma - U$. Then, we have subtracted 1 to those singular values that are bigger than

1 and we have added 1 to the singular values that are less than 1. Suppose that $\sigma_1 \leqslant k$ and $k > 2$. Repeating this process $k-3$ more times results in a matrix whose (diagonal) entries are only between 0 and 2.

Suppose that $\sigma_1 \leqslant 2$. Then $B \equiv \operatorname{diag}(\sigma_1, \sigma_2)$ can be written as a sum of four or more orthogonal matrices, say $B = \sum_{i=1}^{t} Q_i$. If t is even, set $P_i = Q_i \oplus \left[(-1)^{i+1}\right]$. Let $C = \operatorname{diag}(\sigma_1, \sigma_2, \sigma_3)$ and let $D = \sum_{i=1}^{t} P_i$. Then $C - D = \operatorname{diag}(0, 0, \sigma_3)$ can be written as a sum of four orthogonal matrices. If t is odd and if $\sigma_3 \geqslant 1$, set $P_i = Q_i \oplus \left[(-1)^{i+1}\right]$. Then $C - D = \operatorname{diag}(0, 0, \sigma_3 - 1)$ is a sum of four orthogonal matrices. If t is odd and if $\sigma_3 < 1$, set $P_i = Q_i \oplus \left[(-1)^{i}\right]$. Then $C - D = \operatorname{diag}(0, 0, \sigma_3 + 1)$ is a sum of four orthogonal matrices.

Hence, A can be written as a sum of $k + 6$ or more orthogonal matrices.

2.3. The case $\mathbb{F} = \mathbb{H}$

Let $\mathbb{H}$ be the set of quaternions with real coefficients, that is, let $\mathbb{H} = \{a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k} : a, b, c, d \in \mathbb{R}\}$. Let $M_n(\mathbb{H})$ be the set of n-by-n matrices with entries from $\mathbb{H}$. See [6] for a discussion on $M_n(\mathbb{H})$.

Notice that $\mathcal{U}_1(\mathbb{H}) = \{z \in \mathbb{H} : |z| = 1\}$. Let $z = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k} \in \mathbb{H}$ be given. Then $z = x + y\mathbf{j}$, where $x = a + b\mathbf{i} \in \mathbb{C}$ and $y = c + d\mathbf{i} \in \mathbb{C}$. Now, x and y can be written as a sum of elements of $\mathcal{U}_1(\mathbb{C})$. Note that if $p \in \mathcal{U}_1(\mathbb{C})$, then p and $p\mathbf{j}$ are elements of $\mathcal{U}_1(\mathbb{H})$.

One checks that $\mathcal{U}_n(\mathbb{H})$ forms a group under multiplication, so that Lemma 1 applies. Let $n \geqslant 2$. Then, xI and $yI \in M_n(\mathbb{C})$ can be written as a sum of matrices from $\mathcal{U}_n(\mathbb{C})$. Now, if $U \in \mathcal{U}_n(\mathbb{C})$, then U and $U\mathbf{j} \in \mathcal{U}_n(\mathbb{H})$. Hence, zI can be written as a sum of matrices in $\mathcal{U}_n(\mathbb{H})$. Following the proof of Lemma 2, this shows that αE_{11} can be written as a sum of matrices from $\mathcal{U}_n(\mathbb{H})$. Now, for permutation matrices $P, Q \in \mathcal{U}_n(\mathbb{H})$, notice that $P(\alpha E_{11})Q = \alpha P E_{11} Q$, as $\alpha \in \mathbb{H}$ commutes with real matrices. Thus, every $A \in M_n(\mathbb{H})$ can be written as a sum of matrices from $\mathcal{U}_n(\mathbb{H})$.

Let $A \in M_n(\mathbb{H})$ be given, and write $A = U\Sigma V$. Suppose that $\sigma_1(A) \leqslant k$, and suppose that $k \geqslant 2$. Then Σ can be written as a sum of k unitary matrices in $M_n(\mathbb{C})$. Hence, A can be written as a sum of k matrices from $\mathcal{U}_n(\mathbb{H})$.

Theorem 16. *Let n be a given positive integer. Let $A \in M_n(\mathbb{H})$ be given. Then A can be written as a sum of matrices from $\mathcal{U}_n(\mathbb{H})$. Moreover, if $k \geqslant 2$ is an integer such that $\sigma_1(A) \leqslant k$, then A can be written as a sum of k matrices from $\mathcal{U}_n(\mathbb{H})$.*

Let $A, B \in M_n(\mathbb{H})$ be given. We say that A is *orthogonal* if $AA^T = I$. Notice that $(AB)^T$ is not necessarily equal to $B^T A^T$. When $n = 1$, the equality fails to hold because multiplication of quaternions is not commutative. Let $C = \begin{bmatrix} \sqrt{2} & \mathbf{i} \\ \mathbf{i} & -\sqrt{2} \end{bmatrix}$ and let $D = \begin{bmatrix} \sqrt{2} & \mathbf{j} \\ \mathbf{j} & -\sqrt{2} \end{bmatrix}$. One checks that both C and D are orthogonal. However, $CD = \begin{bmatrix} 2 + \mathbf{k} & \sqrt{2}\mathbf{j} - \sqrt{2}\mathbf{i} \\ \sqrt{2}\mathbf{i} - \sqrt{2}\mathbf{j} & 2 + \mathbf{k} \end{bmatrix}$ is not orthogonal, that is, the set of orthogonal matrices in $M_2(\mathbb{H})$ does not form a group under multiplication. Let $n \geqslant 2$. Set $E = C \oplus I_{n-2}$ and set $F = D \oplus I_{n-2}$. Then, $E, F \in \mathcal{O}_n(\mathbb{H})$ but $EF \notin \mathcal{O}_n(\mathbb{H})$. Hence, $\mathcal{O}_n(\mathbb{H})$ does not form a group under multiplication.

Let an integer $n \geqslant 2$ and let $X, Y \in M_n(\mathbb{R})$. Then $X + Y\mathbf{i} \in M_n(\mathbb{C})$ can be written as a sum of matrices from $\mathcal{O}_n(\mathbb{C})$. Similarly, $X + Y\mathbf{j}$ can be written as a sum of matrices from $\mathcal{O}_n(\mathbb{R} + \mathbb{R}\mathbf{j})$ and $X + Y\mathbf{k}$ can be written as a sum of matrices from $\mathcal{O}_n(\mathbb{R} + \mathbb{R}\mathbf{k})$. Hence, every $A \in M_n(\mathbb{H})$ can be written as a sum of matrices from $\mathcal{O}_n(\mathbb{H})$.

Theorem 17. *Let $n \geqslant 2$ be a given integer. Let $A \in M_n(\mathbb{H})$ be given. Then A can be written as a sum of matrices from $\mathcal{O}_n(\mathbb{H})$.*

References

[1] R.A. Horn, C.R. Johnson, Matrix Analysis, Cambridge University Press, New York, NY, 1985.
[2] R.A. Horn, C.R. Johnson, Topics in Matrix Analysis, Cambridge University Press, New York, NY, 1991.
[3] R.A. Horn, D.I. Merino, Contragredient equivalence: a canonical form and some applications, Linear Algebra Appl. 214 (1995) 43–92.
[4] C.K. Li, Y.T. Poon, N.S. Sze, Eigenvalues of the sum of matrices from unitary similarity orbits, SIAM J. Matrix Anal. Appl. 30 (2) (2008) 560–581.
[5] F.Y. Wu, Additive combinations of special operators, Functional Analysis and Operator Theory Banach Center Publication, vol. 30, Institute of Mathematics, Polish Academy of Sciences, Warszawa, 1994, pp. 337–361.
[6] F. Zhang, Quaternions and matrices of quaternions, Linear Algebra Appl. 251 (1997) 21–57.